Pattern Recognition xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Group-aware deep feature learning for facial age estimation

Hao Liu[a,b], Jiwen Lu[a,b,*], Jianjiang Feng[a,b], Jie Zhou[a,b]

[a] *Department of Automation, Tsinghua University, Beijing 100084, China*
[b] *Tsinghua National Laboratory for Information Science and Technology, Beijing 10084, China*

ARTICLE INFO

Keywords:
Facial age estimation
Deep learning
Feature learning
Biometrics

ABSTRACT

In this paper, we propose a group-aware deep feature learning (GA-DFL) approach for facial age estimation. Unlike most existing methods which utilize hand-crafted descriptors for face representation, our GA-DFL method learns a discriminative feature descriptor per image directly from raw pixels for face representation under the deep convolutional neural networks framework. Motivated by the fact that age labels are chronologically correlated and the facial aging datasets are usually lack of labeled data for each person in a long range of ages, we split ordinal ages into a set of discrete groups and learn deep feature transformations across age groups to project each face pair into the new feature space, where the intra-group variances of positive face pairs from the training set are minimized and the inter-group variances of negative face pairs are maximized, simultaneously. Moreover, we employ an overlapped coupled learning method to exploit the smoothness for adjacent age groups. To further enhance the discriminative capacity of face representation, we design a multi-path CNN approach to integrate the complementary information from multi-scale perspectives. Experimental results show that our approach achieves very competitive performance compared with most state-of-the-arts on three public face aging datasets that were captured under both controlled and uncontrolled environments.

1. Introduction

Facial age estimation attempts to predict the real age value or age group based on facial images, which has widely potential applications such as facial bio-metrics, human-computer interaction, social media analysis and entertainments [1–4]. While extensive efforts have been devoted, facial age estimation still remains a challenging problem due to two aspects: 1) lack of sufficient training data where each person should contain multiple images in a wide range of ages, 2) large variations such as lighting, occlusion and cluttered background of face images which were usually captured in wild conditions.

Most existing facial age estimation systems usually consist of two key modules: face representation and age estimation. Representative face representation approaches include holistic subspace features [5,6], active appearance model (AAM) [7], Gabor wavelets [7], local binary pattern (LBP) [8] and bio-inspired feature (BIF) [9]. Having obtained face representations, age estimation can be addressed as a classification or regression problem [9–11]. However, the face representations employed most existing methods are hand-crafted, which requires strong prior knowledge to engineer it by hand. To address this problem, learning-based feature representation methods [5,12,13,3] have been made to learn discriminative feature representation directly from raw pixels. However, these methods aim to learn linear feature filters to project face images into another feature space such that they may not be powerful enough to exploit the nonlinear relationship of data. To address this nonlinear issue, deep learning-based methods have been adopted to learn a series of nonlinear mapping functions between face image and age label [14–16,16–18]. Unfortunately, these deep models cannot explicitly achieve the ordinal relationship among the chronological ages, which are still far from the practical satisfactory in most cases because they usually encounter unbalanced and insufficient training data for each age label.

Notice that age labels are chronologically correlated, so that it is desirable to employ nonlinear discriminative methods to exploit the correlated order information from facing images. Unlike existing deep learning-based facial age estimation methods that ignored the ordinal information of face aging data, we proposed a group-aware deep feature learning method (GA-DFL) under deep convolutional neural networks (CNN), by learning discriminative face representations directly from image pixels and exploiting the aging order information. Since facial aging datasets usually lack of face images from the same person covering a wide range of ages, our proposed GA-DFL first separates the chronological aging progress into several overlapped groups and then learns a series of hierarchical nonlinear mapping

* Corresponding author at: Department of Automation, Tsinghua University, Beijing 100084, China.
E-mail address: elujiwen@gmail.com (J. Lu).

http://dx.doi.org/10.1016/j.patcog.2016.10.026
Received 13 July 2016; Received in revised form 29 September 2016; Accepted 20 October 2016
Available online xxxx
0031-3203/

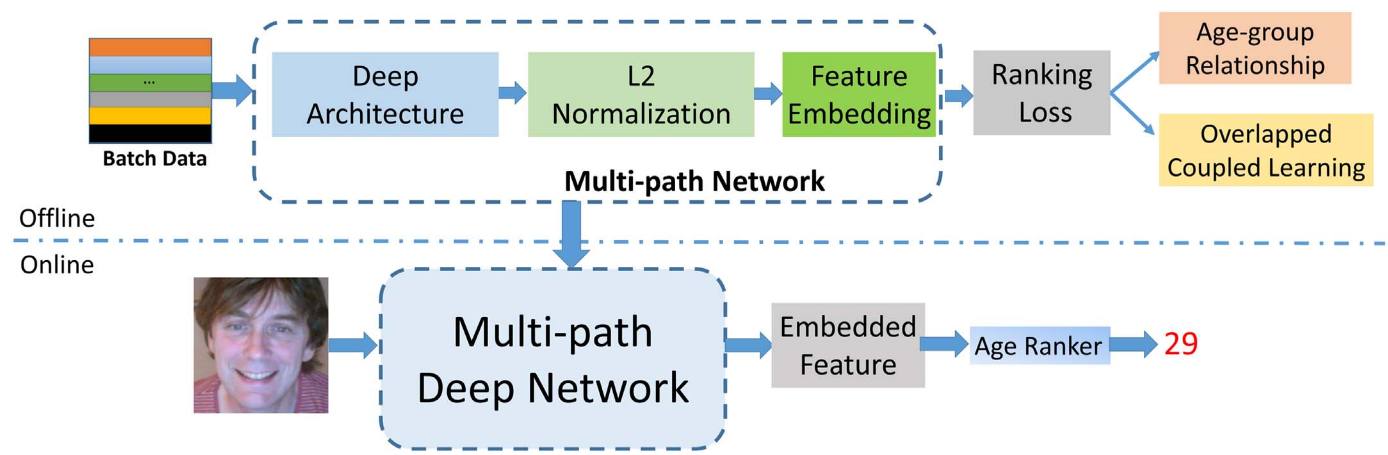

Fig. 1. The pipeline of the proposed facial age estimation approach. During the offline phase, we enforce two criterions on modeling aging progress to learn face representation: 1) the inter-group variances are maximized while the intra-group variances are minimized. 2) the separated age groups should be smoothed on the age-group specific overlaps. Then the parameters of the designed network are optimized by back-propagation. During online phase, we feed face image into the network to obtain face representation and the final age label is performed by an age ranker.

functions to project raw pixel values into another common feature space, so that face pairs in the same age groups are projected as close as possible while those in different age groups are projected as far as possible. Moreover, we link every discrete groups by overlapping structures and develop an overlapped coupled learning method, which aims to smooth the age differences lying on the overlaps of the adjacent age groups. We also propose a multi-path CNN architecture to enhance the capacity of feature representation to integrate complementary information from multiple scales to improve the performance. Fig. 1 illustrates the main procedure of our proposed approach. To evaluate the effectiveness of our proposed GA-DFL, we conducted experiments on three widely used facial age estimation datasets that were captured in both constrained and unconstrained environments. Experimental results show that our proposed GA-DFL obtains superior performance compared with most state-of-the-art facial age estimation methods.

The contributions of this work are summarized as follows:

(1) We develop a deep feature learning method to discriminatively learn a face representation directly from raw pixels. With the learned nonlinear filters, the chronological age information can be well exploited with a perspective of age groups in the obtained face descriptor.
(2) We propose an overlapped coupled learning method to achieve the smoothness on the neighboring age groups. With this learning strategy, the age difference information on the age-group specific overlaps can be well measured.
(3) We employ a multi-path deep CNN architecture to integrate multiple scale information into the learned face presentation.

The rest of this paper is organized as follows: Section 2 reviews some backgrounds. Section 3 details the proposed GA-DFL method. Section 4 provides the experimental results and Section 5 concludes this paper.

2. Related work

In this section, we briefly review two related topics: facial age estimation and deep learning.

2.1. Facial age estimation

Numerous facial age estimation methods [19–22,12,23–27] have been proposed over the past two decades. As one of the earliest studies, Lanitis et al. [25] applied a quadratic function to predict facial age. Thereafter, several works [28,23] were proposed to incorporated with correlated age labels to model practical human aging progress with different degrees of improvements. In particular, Chang et al. [28] presented an ordinal hyperplane ranking method (OHRank) which divided age classification as a series of sub-problems of binary classification. Geng et al. [23] proposed a label distribution learning approach to model the relationship between face images and age labels. Besides, Guo and Mu [29] showed human gender and race are used to exploit the complementary information for age estimation. However, most of these methods utilize hand-crafted features, which require strong prior knowledge by-hand and usually encounters time-consuming. To address this, several studies have been made to learn a discriminative face representation by using advanced feature learning approaches [30,24,3]. For example, Guo et al. [30] proposed a holistic feature learning approach utilizing manifold learning technique. Lu et al. [3] employed a local binary feature learning method to learn a face descriptor robust to local illumination, which has achieved considerable performances on facial age estimation. Nevertheless, these methods focus on learning linear filters so that they are not powerful enough to describe the age-informative facial appearances because there are large variances on collected face data due to scaling, occlusion and cluttered background especially captured in wild conditions. In contrast to these previous works, we propose a deep learning method from a perspective of feature learning with a feed-forward neural networks to exploit the nonlinear relationship of data.

2.2. Deep learning

In the recent literature, deep learning has received much attention in the research field of machine learning and computer vision due to its superior performance in learning a series of nonlinear feature mapping functions directly from raw pixels. A number of deep learning approaches such as restricted Boltzmann machine (RBM) [31], stacked denoising auto-encoder (SDAE) [32], deep convolutional neural networks (CNN) [33] have been successfully employed in many visual analysis tasks such as handwritten digit recognition [34], object detection [35], visual tracking [36] and scene labeling [37]. More recently, deep learning methods have been applied to face analysis tasks including face detection [38], face alignment [39] and face recognition [40,41]. Specifically, Zhang et al. [39] proposed a deep learning method with stacked auto-encoder networks to estimate facial landmarks in a coarse-to-fine manner, Sun et al. [40] developed DeepID2 network to reduce the personalized inter-covariance joint by identification and verification, and Parkhi et al. [41] employed a very deep architecture VGG-16 Face Net pre-trained by a large scale face dataset to perform face recognition.